Journal of Magnetic Resonance 210 (2011) 146–150

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

Communication

Use of continuous optimization methods to find carbon links in 2D INADEQUATE spectra

Christopher Kumar Anand [a], Alex D. Bain [b], Sean C. Watson [a,*]

[a] Department of Computing and Software, McMaster University, Hamilton, ON, Canada
[b] Department of Chemistry and Chemical Biology, McMaster University, Hamilton, ON, Canada

ARTICLE INFO

Article history:
Received 10 November 2010
Revised 8 February 2011
Available online 24 February 2011

Keywords:
NMR
INADEQUATE
Optimization
Image regularization
Gauss-Seidel

ABSTRACT

The 2-D INADEQUATE experiment is a useful experiment for determining carbon structures of organic molecules, which is known for having low signal-to-noise ratios. A non-linear optimization method for solving low-signal spectra resulting from this experiment is introduced to compensate. The method relies on the peak locations defined by the INADEQUATE experiment to create boxes around these areas and measure the signal in each. By measuring pairs of these boxes and applying penalty functions that represent a priori information, we are able to quickly and reliably solve spectra with an acquisition time approximately a quarter of that required by traditional methods. Examples are shown using the spectrum of sucrose.

1. Introduction

The INADEQUATE experiment [1–5] is a 2-D NMR [6] technique used to find carbon-carbon bonds, with the full spectrum giving complete information on the carbon skeleton of the sample molecule [7–11]. However, the low sensitivity of this type of experiment, forcing long experiment times is well documented [12–17].

In this communication, we propose a method for processing the data from an INADEQUATE experiment. It is based on previous image denoising and regularization methods [18–21] to allow the use of this experiment at much lower signal-to-noise ratios than are possible using traditional Fourier transform methods. Since there is much a priori information available and the spectral components are very simple, the INADEQUATE experiment is particularly amenable to more sophisticated processing compared to other 2D experiments. We show an example of a speed-up of approximately four times in this communication, and based on early experimentation, these results are typical.

Before we start an INADEQUATE experiment, we know the single quantum frequencies (hence, all possible double-quantum frequencies); it is their assignment that is unclear. The correlations in the INADEQUATE plots all appear as simple AB spectra. Since both parts of a correlation come from the same double-quantum coherence, the two parts of the correlation should be mirror images [22], although pathological offset effects [23] may slightly distort the symmetry. These key facts allow us to build penalty functions who guide the optimization toward the global minimum. The process makes use of image regularization techniques to smooth the lineshapes as the optimization progresses. The combination of these allows us to find a complete set of carbon bonds from an INADEQUATE spectrum too noisy for traditional methods to be applied, significantly reducing the required experiment time. While the examples shown in this paper use unprocessed data, running the Fourier transforms in MATLAB, this method will work equally well with processed data as the processing will not change any of these assumptions made. Many existing techniques developed for INADEQUATE spectra rely on complex line-shape fitting and large numbers of assumptions on the form of the experimental results [24–31]. Our technique is based on the magnitude of the signal, making this technique insensitive to phase errors in the experiments and uses comparably few assumptions about the form of the experiment in order to simplify our model.

2. Optimization approach

Our approach is based, primarily, on the signal intensity present in a set of pre-defined locations based on the ^{13}C chemical shifts. In the INADEQUATE experiment, these locations are well-defined by the 1-D carbon spectrum. With a list of chemical shifts for all the carbons in the molecule we are testing, we are able to locate all possible peak locations and know which locations of peaks could indicate what bond. We draw boxes around all of these locations, large enough to contain the full peak and account for small shifts of location caused by the coupling constants and isotope effects

* Corresponding author. Address: Department of Computing and Software, McMaster University, 1280 Main St. W., Hamilton, Ontario, Canada L8S 4K1.
E-mail address: watsos@mcmaster.ca (S.C. Watson).

1090-7807/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2011.02.018

(testing shows 32 points along the F_2 axis to be sufficient for an experiment using 4k points). In order to provide more intelligent estimates to compensate for high noise levels, we provide several penalties based on a priori knowledge of the experiment and basic chemistry. We penalize large differences between signal level in paired boxes as the INADEQUATE experiment will typically have equal signal in each of the two correlated boxes and we force the number of bonds to any single carbon to be between one and four.

The method presented allows and requires some user input. We require an estimate of the level of connectivity in the molecule, as this controls the number of bonds we reject or accept as the algorithm runs. We allow an initial guess of the bonds that will be present, in small molecules this is not required, but in large molecules any prior knowledge will greatly improve the run-time of the method.

2.1. Variables

The two sets of variables we use are intended to meet the goals of the optimization:

p: to determine the bonds present in a molecule, given a 2-D INADEQUATE spectrum; and
$\mathscr{S}$: to create a cleaner image of the given spectrum, clearly showing the locations of the peaks and bonds determined.

The column vector, p, contains one entry for each bond (a total of n choose 2 entries, where n is the number of carbons present). Each entry contains a value ranging from 0 to 1 that represents the probability of its associated bond being present in the molecule. The entry of p relating to a specific bond, between carbons i and j, is given by the notation p_{ij}. This notation does not indicate the index of an entry, but rather indicates the contents by noting the two carbons being referenced. We use this non-standard convention in order to simplify the written problem. The entries in this vector are ordered with respect to the bonds they represent. For each carbon i, there is an ordered list of bonds from $i \leftrightarrow i+1$ to $i \leftrightarrow n$ (where $\leftrightarrow$ means "bonded to") where i ranges between 1 and $n-1$. The carbons themselves are numbered by their chemical shifts in increasing order.

The variable $\mathscr{S}$ is a sparse array with dimensions equal to those of the original Fourier transformed spectrum. $\mathscr{S}$ has the important structure of being zero-valued outside of the pre-determined boxes that may contain peaks (determined by the chemical shifts of the potentially bonded carbons), meaning that we eliminate noise from all regions that we know cannot contain a correlation. Each of these boxes is made up of a series of entries, referred to as s_k. This structure minimizes the number of variables we need to solve in order to create a new image of the spectrum without a loss of data. A small scale example of this type of structure is shown in Eq. (1). In order to simplify several later equations, we use the notation $\mathscr{S}_{ij}$ to refer to a sub-array of $\mathscr{S}$ corresponding to the box centred at the location $(\omega_i,\omega_i+\omega_j)$, where we could see one of the peaks indicating a bond of i to j.

$$\begin{bmatrix} 0 & s_1 & s_2 & s_3 & 0 & 0 & 0 & s_4 & s_5 & s_6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ s_7 & s_8 & s_9 & 0 & 0 & s_{10} & s_{11} & s_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{1}$$

In the example in Eq. (1), the variables s_1, s_2,s_3 correspond to one box (if this represents a bond between carbons 1 and 2 we refer to it by $\mathscr{S}_{12}$). In a real spectrum we use wider boxes with many more variables per box.

2.2. Model

Our optimization problem minimizes the sum of the terms given in (2)–(7) and is constrained by (8) and (9). The solution of this minimization problem gives us both desired results; a low-noise image of the spectrum and a list of carbon bonds present in the molecule.

$$\begin{aligned} \min \quad & \|m-\mathscr{S}\|^2 && (2)\\ & +\lambda_1\|\delta_x\mathscr{S}\|^2 && (3)\\ & +\lambda_2\sum_{ij}(1-p_{ij})^2\|\mathscr{S}_{ij}+\mathscr{S}_{ji}\|^2 && (4)\\ & +\lambda_3\sum_{ij}(\|\mathscr{S}_{ij}\|^2-\|\mathscr{S}_{ji}\|^2)^2 && (5)\\ & +\mu_1\sum_i\left(2-\sum_{j\neq i}p_{ij}\right)^4 && (6)\\ & +\mu_2\sum_{ij}p_{ij} && (7)\\ \text{s.t. } & p_{ij}\geqslant 0 && (8)\\ & p_{ij}\leqslant 1 && (9) \end{aligned}$$

Term (2) attempts to minimize the difference between the two-dimensional Fourier transform of the measured spectrum (m) and the optimized image of the spectrum ($\mathscr{S}$). This gives us a final image whose peaks are located and shaped in the same way as those in the original spectrum as their size changes. This term also tends to minimize the difference in signal level between the two arrays, meaning that we redistribute the signal of the spectrum rather than reducing or increasing it as we modify the values in the boxes of $\mathscr{S}$.

For term (3), we define

$$\|\delta_x\mathscr{S}\|^2 = \sum_{\substack{i \\ \text{s.t. } s_i \text{ and } s_{i+1} \text{ are horizontally} \\ \text{adjacent in the same box}}} (s_i - s_{i+1})^2,$$

as a short form for the L_2 difference norm in the x (F_2) dimension, taking into account the sparsity of $\mathscr{S}$. This will tend to minimize the differences between adjacent points. Minimizing these differences promotes smoothness of the lines in this direction without trying to fit the peak to a particular line shape, which would require the incorporation of other variables. We do not regularize in the F_1 dimension as we do not generally expect smoothness in this direction.

Term (4) penalizes signal in the boxes of $\mathscr{S}$ if they are considered unlikely to correspond to a bond and penalizes the values in p if they correspond to low-signal areas of $\mathscr{S}$. These values will be increased in the opposite scenario. This promotes increasing

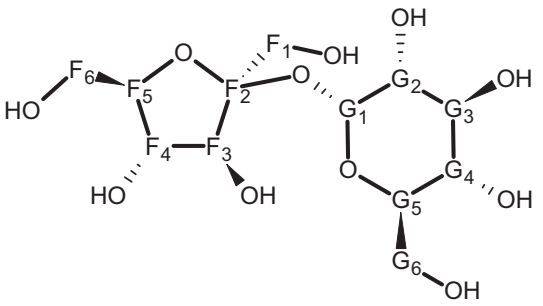


Fig. 1. The structure of sucrose, showing the six-membered glucose ring(G) and five-membered fructose ring(F) with numbered carbons.

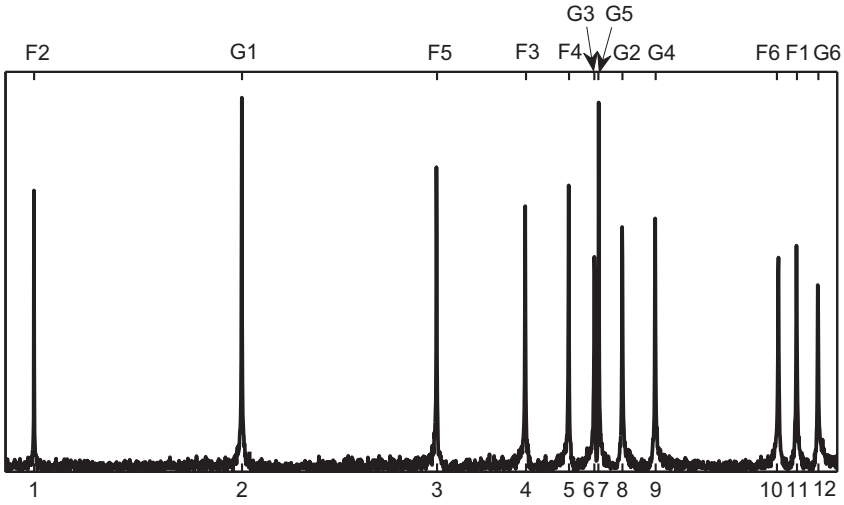


Fig. 2. The numbering of the sucrose peaks with the structural numbering on the top and the numbering used in our method on the bottom.

signal in areas of likely peaks and a decrease of signal in areas of unlikely peaks. In the end, signal will be completely removed from locations we have decided will not contain peaks and will be concentrated in the areas of high certainty peaks.

Term (5) penalizes the signal in two paired boxes of $\mathscr{S}$ if the two boxes have a large difference in signal level. This means that a correlation will become less likely if its two related areas have very different signal levels.

Term (6) is a quartic penalty function designed to keep the total number of bonds for a single carbon between one and four with two bonds being the most likely. The term achieves this by taking each carbon i and summing over the bond probability between i and every other carbon, j to get the total value of the bonds to i. This sum is then subtracted from 2 to shift the center of the quartic to +2 and the result of this is put to the fourth power to generate our function. This function was chosen because it remains fairly flat in the center and sharply rises afterward meaning that two bonds to a carbon is not penalized at all, one or three bonds are lightly penalized and outside that specified range, the penalty is strong.

Term (7) penalizes the sum of the probabilities in p, or the number of likely carbon bonds found. This function simply adds together all the bond probability values and minimizes this value. This is necessary because we do not have an a priori estimate for the peak heights, so we are only penalizing peaks which we do not believe correspond to bonds and large differences in paired peak heights. We would therefore expect a minimum in the objective function when many bonds are predicted, even ones with very low intensity peaks. This penalty prevents that non-descriptive minimum from occurring. In practice, investigators will have a good estimate of the total number of bonds in an unknown, or partially known, molecule and this information would be incorporated into this penalty.

2.3. Implementation

The problem is dominated by bi-quadratic terms; being quadratic in both $\mathscr{S}$ and p separately. This fact suggests that the problem can be split into two parts and solved using an alternating Gauss–Seidel approach to solve alternatively for $\mathscr{S}$ and for p.

In the first, unconstrained, problem (10) we solve for $\mathscr{S}$. The second part (11) is a constrained problem that estimates the likelihoods of all possible bonds, thereby solving for p. The result of this split was a large increase in the speed of the solution. We implement this problem in a MATLAB program[1] using standard routines included in the Mathworks optimization toolbox. We load the raw time series data into MATLAB using matNMR[32] and we run a standard MATLAB Fourier transform before normalizing the spectrum to make the largest value 1 in order to limit the size of the necessary constants. In initial tests using the sucrose spectra at a size of 128×4096, the solution using the single (combined) problem took between 10 and 12 h, while the split problem took between 2 and 10 min to reach comparable solutions.

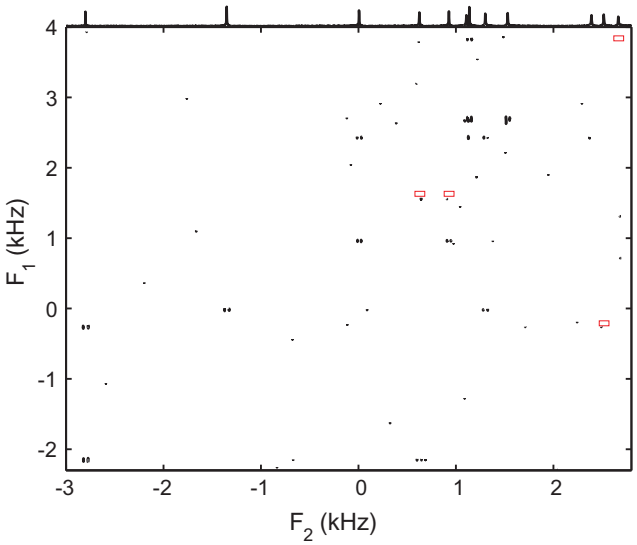


Fig. 3. Averaged Fourier transform of all eight collected data sets. Here we can see most of the structure, but several bonds remain difficult to detect, most notably: $F2 \leftrightarrow F1$, $F3 \leftrightarrow F4$ and $G5 \leftrightarrow G6$. Red boxes indicate locations at which we should see peaks related to these bonds, but cannot.

$$\min_{\mathscr{S}} \ \|m - \mathscr{S}\|^2 + \lambda_1 \|\delta_x \mathscr{S}\|^2 + \lambda_2 \sum_{ij} (1 - p_{ij})^2 \|\mathscr{S}_{ij} + \mathscr{S}_{ji}\|^2 + \lambda_3 \sum_{ij} (\|\mathscr{S}_{ij}\|^2 - \|\mathscr{S}_{ji}\|^2)^2 \tag{10}$$

[1] The program is available on request.

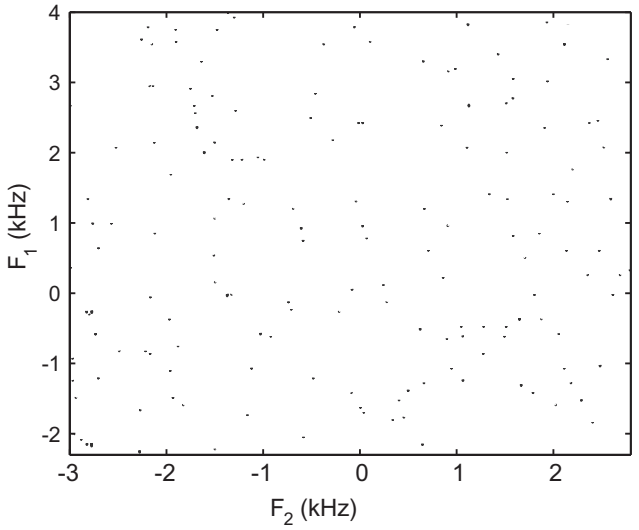
The first (unconstrained) problem results in the redistribution of signal from the original spectrum which causes an increase of peak area in the possible bond locations and a decrease of peak area across the rest of the spectrum. This means that we will achieve a clearer image of the spectrum.

$$\begin{aligned} &\min_{p_{ij}} \sum_{ij}(1-p_{ij})^2\|\mathscr{S}_{ij}+\mathscr{S}_{ji}\|^2+\mu_1\sum_i\left(2-\sum_{j\neq i}p_{ij}\right)^4+\mu_2\sum_{ij}p_{ij}\\ &\text{s.t. } p_{ij}\geqslant 0\\ &\quad p_{ij}\leqslant 1 \end{aligned} \tag{11}$$

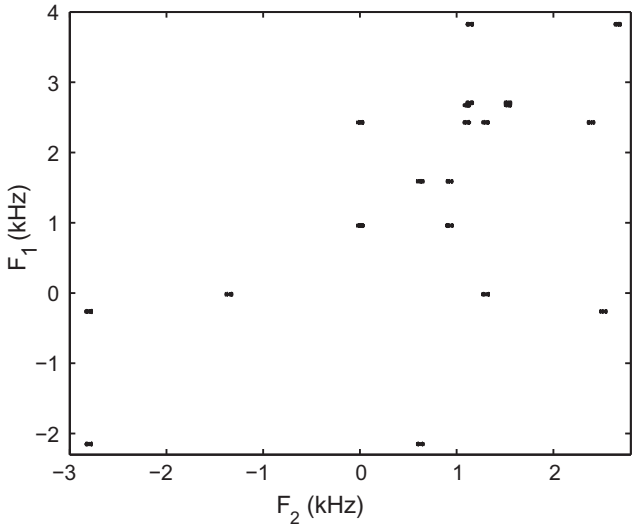
The second (constrained) problem estimates the bond probabilities.

This splitting is the most natural and it results in two subproblems with desirable performance properties. The original problem is a large, non-quadratic, constrained problem. When we split it into two however, the first subproblem is a large, but mostly quadratic problem and the second, while non-quadratic and constrained, involves many fewer variables than the original problem. This relationship is maintained for all sizes of molecules.

The penalty parameters λ_i and μ_i are scaling factors for each of the terms we wish to minimize. Generally these constants do not need to be modified from default values where we assign: $\lambda_1 = 10$, $\lambda_2 = 15$, $\lambda_3 = 1$, $\mu_1 = 4$. However, the experiment is rather sensitive to the value of μ_2, which represents the degree of connectivity, and this value must be changed according to the amount of signal and the connectivity of a molecule.



(a) Unmodified FT of two averaged data sets



(b) Results of running our method on (a)

Fig. 4. We can see here that running our method on two averaged data sets gives us better information than the average of all eight using traditional methods, constituting a decrease in experiment time of a factor of more than four.

3. Results

The experiment we performed was with a concentrated sucrose (see Figs. 1 and 2) solution (approximately 80mg of sucrose in 0.5 mL), run on a Bruker AV 500MHz spectrometer equipped with a 5 mm room-temperature inverse-geometry probe. The pulse sequence we used was the standard INADEQUATE experiment including gradients to enhance pathway selection. Instead of running a single experiment overnight, we run several smaller experiments sequentially in order to help prevent any computer or spectrometer errors from affecting the entire dataset. Each of these experiments result in what we term a block; when averaged, these blocks give the results of the full experimental time. We gathered a total of eight blocks of data, each measuring 512 points in T_1 by 4096 in T_2 using 32 scans with a scan delay of one second and a 90° pulse with width 15 μs taking slightly over 6 h to run. The optimization is run using the optimization toolbox of MATLAB 7.9.0 installed on a department server using a dual-core AMD Opteron processor at 2.6 GHz. For small problems like this (under 20 or so carbons), the MATLAB solver is sufficiently fast, however for much larger problems as we expect to encounter, we will need to explore options for faster solvers.

In Fig. 3, we see the Fourier transform of the averaged value of all eight blocks. We can visually determine seven of the ten bonds reliably, but we have two bonds where only one of the two peaks are visible and one where neither is visible. In Fig. 4 we see a Fourier transform of the averaged value of only two blocks with the result of running our method on that data. We see that even with only a quarter of the data, the final result of our method is much clearer than what can be seen in Fig. 3. Using two of the eight blocks was the minimum we could reliably use, but there existed a few combinations (all including one specific block) of two blocks that we could not solve. The results for all other pairs of blocks look highly similar to the results shown here.

4. Conclusion

We have a shown a method of solving for carbon bonds from INADEQUATE spectra by using image regularization and optimization techniques. Despite the limited amount of testing so far, the early results of this method of solving for carbon bonds from INADEQUATE spectra are promising. We have shown that we can reduce the experiment time required to solve from the spectrum of sucrose from over 48 h using traditional methods to around 12 using our method. We plan to expand our testing to a wide variety of molecules as soon as possible and we expect similar speed-ups in all cases.

References

[1] A. Bax, R. Freeman, Investigation of 13C–13C couplings in natural-abundance samples: the strong coupling case, Journal of Magnetic Resonance 41 (1980) 507–511.
[2] A. Bax, R. Freeman, S.P. Kempsell, Investigation of 13C–13C long-range couplings in natural-abundance samples, Journal of Magnetic Resonance 41 (1980) 349–353.

[3] A. Bax, R. Freeman, S.P. Kempsell, Natural abundance 13C–13C coupling observed via double-quantum coherence, Journal of the American Chemistry Society 102 (1980) 4849–4851.
[4] A. Bax, R. Freeman, T.A. Frenkiel, M.H. Levitt, Assignment of carbon-13 NMR spectra via double quantum coherence, Journal of Magnetic Resonance 43 (1981) 478–483.
[5] A. Bax, T.H. Mareci, Practical aspects of carbon-13 double quantum NMR, Journal of Magnetic Resonance (1969) 53 (1983) 360–363.
[6] R.R. Ernst, G. Bodenhausen, A. Wokaun, Principles of Nuclear Magnetic Resonance in One and Two Dimensions, Clarendon, Oxford, 1987.
[7] A. Bax, R. Freeman, T.A. Frenkiel, An NMR technique for tracing out the carbon skeleton of an organic molecule, Journal of the American Chemistry Society 103 (1981) 2102–2104.
[8] R. Freeman, T.A. Frenkiel, M.B. Rubin, Structure of a photodimer determined by natural-abundance 13C–13C coupling, Journal of the American Chemistry Society 104 (1982) 5545–5547.
[9] A. Neszmelyi, G. Lukacs, Natural abundance one-bond C-13–C-13 coupling constants in monosaccharide derivatives and in sucrose, Journal of the American Chemistry Society 104 (1982) 5342–5346.
[10] J. Lambert, H.J. Kuhn, J. Buddrus, Complete identification of C–C connectivities by two-dimensional INADEQUATE NMR spectroscopy with composite pulses, Angewandte Chemie International Edition English 28 (1989) 738–740.
[11] Z. Zhou, J.C. Stevens, J. Klosin, R. Kümmerle, X. Qui, D. Redwine, R. Cong, A. Taha, J. Mason, B. Winniford, P. Chauvel, N. Montanez, NMR study of isolated 2, 1-inverse insertion in isotactic polypropylene, Macromolecules 42 (2009) 2291–2294.
[12] J. Buddrus, H. Bauer, Direct identification of the carbon skeleton of organic compounds using double quantum coherence ^{13}C-NMR spectroscopy. The INADEQUATE pulse sequence, Angewandte Chemie International Edition in English 26 (1987) 625–642.
[13] B.H. Oh, W.M. Westler, P. Darba, J.L. Markley, Protein carbon-13 spin systems by a single two-dimensional nuclear magnetic resonance experiment, Science 240 (1988) 908–911.
[14] D.W. Bohman, Z. Li, D. Li, N.L. Owen, The analysis of mixtures using the 2-D INADEQUATE NMR technique, Journal of Molecular Structure 480–481 (1999) 125–132.
[15] O.W. Sørensen, R. Freeman, T. Frenkiel, T.H. Mareci, R. Schuck, Observation of 13C–13C couplings with enhanced sensitivity, Journal of Magnetic Resonance (1969) 46 (1982) 180–184.
[16] P.J. Keller, K.E. Vogele, Sensitivity enhancement of INADEQUATE by proton monitoring, Journal of Magnetic Resonance (1969) 68 (1986) 389–392.
[17] I.S. Podkorytov, Seven-pulse sequence DEPT-INADEQUATE, Journal of Magnetic Resonance (1969) 89(1990) 129–132.
[18] H. Andrews, B. Hunt, Digital Image Resoration, Prentice-Hall, 1977.
[19] M. Bertero, P. Boccacci, Introduction to Inverse Problems in Imaging, IOP Publishing Ltd., 1998.
[20] J.M. Bardsley, C.R. Vogel, A nonnegatively constrained convex programming method for image reconstruction, SIAM Journal on Scientific Computing 25 (2003) 1326–1343.
[21] P.C. Hansen, J.G. Nagy, D.P. O'Leary, Deblurring Images: Matrices, Spectra and Filtering, SIAM, 2006.
[22] T. Nakazawa, H. Sengstschmid, R. Freeman, Enhancement of INADEQUATE spectra according to symmetry criteria, Journal of Magnetic Resonance 120 (1996) 269–273.
[23] A.D. Bain, D.W. Hughes, C.K. Anand, Z. Nie, V.J. Robertson, Problems, artifacts and solutions in the INADEQUATE NMR experiment, Magnetic Resonance in Chemistry 48 (2010) 630–641.
[24] S.W. Sparks, P.D. Ellis, DEPT polarization transfer for the INADEQUATE experiment, Journal of Magnetic Resonance (1969) 62(1985) 1–11.
[25] K.E. Kövér, D. Uhrin, T. Liptaj, G. Batta, Easy implementation of selective INADEQUATE and H–C–C relay experiments using heteronuclear chemical shift filtering, Magnetic Resonance in Chemistry 30 (1992) 68–72.
[26] J. Lambert, J. Buddrus, Sensitivity enhancement of two-dimensional 13C, 13C-INADEQUATE spectroscopy by considering symmetry and isotopic shifts, Journal of Magnetic Resonance, Series A 101 (1993) 307–312.
[27] R. Dunkel, C.L. Mayne, J. Curtis, R.J. Pugmire, D.M. Grant, Computerized analysis of 2D INADEQUATE spectra, Journal of Magnetic Resonance (1969) 90(1990) 290–302.
[28] R. Dunkel, C.L. Mayne, R.J. Pugmire, D.M. Grant, Improvements in the computerized analysis of 2D INADEQUATE spectra, Analytical Chemistry 64 (1992) 3133–3149.
[29] R. Richarz, W. Amman, T. Wirthlin, Computer-assisted determination of carbon connectivity patterns in organic molecules from natural-abundance $^{1}J_{CC}$ data, Journal of Magnetic Resonance 45 (1981) 270–283.
[30] M.P. Foster, C.L. Mayne, R. Dunkel, R.J. Pugmire, D.M. Grant, J.-M. Kornprobst, J.-F. Verbist, J.-F. Biard, C.M. Ireland, Revised structure of bistramide a(bistratene a): application of a new program for the automated analysis of 2D INADEQUATE spectra, Journal of the American Chemistry Society 114 (1992) 1110–1111.
[31] A.M. Orendt, R. Dunkel, W.J. Horton, R.J. Pugmire, D.M. Grant, Computerized analysis of 2D INADEQUATE spectra to assign chemical shifts in aromatic compounds, Journal of Magnetic Resonance 33 (1995) 803–811.
[32] J.D. van Beek, matNMR: a flexible toolbox for processing, analyzing and visualizing magnetic resonance data in Matlab, Journal of Magnetic Resonance 187 (2007) 19–26.